氏名	スズムラ シンヤ **鈴村 真矢**
学位の種類	博士（工学）
学位記番号	博第1217号
学位授与の日付	2021年3月31日
学位授与の条件	学位規則第４条第１項該当　課程博士
学位論文題目	Statistical inference for feature selection algorithms and how to construct robust learning models （特徴量選択アルゴリズムに関する統計的推論とロバスト学習の構築に向けた研究）
論文審査委員	主査　教授　竹内 一郎 教授　玉木 徹 准教授　烏山 昌幸 教授　池田 和司 （奈良先端科学技術大学院大学）

論文内容の要旨

研究の背景・目的：

この数十年間にわたり，機械学習（例えば深層学習）は強力なデータ解析ツールとして有用性を示しており，その応用先は音声認識，画像認識，自然言語処理など様々である．
他方，機械学習は内部構造がブラックボックス化されることが多く，学習済みモデルの解釈性の向上が求められている．モデルの解釈性は，例えばバイオインフォマティックスの研究領域において非常に重要視されている．この領域では，例えば，ある疾患に作用する新薬開発の研究を考えたとき，どのような患者がどのような薬剤に効果を示すか，効果の大きさはどれくらいか，といった情報をデータ解析結果をもとに定量的に示すことが求められる．また，データ解析結果が実社会に応用されることを想定すると，結果が偶然に得られたものでないという裏付けの定量化も重要となる．上記の理由により，本研究では，機械学習の結果の解釈を統計的なアプローチで定量化することに焦点をおき，その上で実データ解析として有力な機械学習アルゴリズムを提案する．

具体的な研究内容：

本研究は二章で構成される．第一章では，機械学習アルゴリズムにより選択された特徴量の統計的な有用性を定量化するための新しい手法を提案した．

この研究により，レスポンス（例えば患者の薬剤耐性）に関連する特徴量間の組み合わせ（組み合わせの次数は最大無限大まで考慮可能）を抽出し，レスポンスとの関連性を定量化することが可能になった．また，実データ実験において，提案法を用いて HIV-1 低ウイルス薬に関連する遺伝子配列の突然変異の組み合わせを抽出し，得られた結果の統計的な信頼性を定量化した．

第二章では，機械学習の学習データに潜む外れ値（Outlier）が学習結果に与える影響度合いを自動的に制御し，学習結果を安定させるというロバスト学習の新手法を提案した．この研究のモチベーションは，近年，クラウドソーシングなどを利用して比較的容易にデータを取得することができる反面，それらのデータに潜むノイズにより学習結果が不安定になるという問題点を克服することである．提案法は，あるデータがどの程度異常であるかを自動的に判定しその影響度を調節することで学習結果を安定させる．影響度の調節はアニーリング（焼きまなし）法により実現されるが，提案法は無限に細かい粒度で（連続的に）温度パラメータを調節することが計算上可能であり，温度パラメータを離散的に調節する従来法と比較して高い汎化性能が得られることが数値実験で示された．また，本研究では提案アルゴリズムによって得られる解の性質を理論的な側面から解析し，アルゴリズムの解釈性の向上を図った．

論文審査結果の要旨

近年、機械学習（例えば深層学習）は強力なデータ解析ツールとして有用性を示しており，その応用先は音声認識，画像認識，自然言語処理など様々な分野に及ぶ．機械学習モデルの多くは内部構造がブラックボックス化されることが多いが，応用分野によっては学習済みモデルの解釈性の向上が求められている．例えば、バイオインフォマティックス分野では，ある疾患に作用する新薬開発の研究を考えたとき，どのような患者がどのような薬剤に効果を示すか，効果の大きさはどれくらいか，といった情報をデータ解析結果をもとに定量的に示すことが求められる．また，データ解析結果が実社会に応用されることを想定すると，結果が偶然に得られたものでないという裏付けの定量化も重要となる．本論文では、機械学習の結果の解釈を統計的なアプローチで定量化することに焦点をおき，その上で実データ解析として有力な機械学習アルゴリズムが提案されている。

第一章では，機械学習アルゴリズムにより選択された特徴量の統計的な有用性を定量化するための新しい手法を提案されている。この研究では，レスポンス（例えば患者の薬剤耐性）に関連する特徴量間の組み合わせ（組み合わせの次数は最大無限大まで考慮可能）を抽出し，レスポンスとの関連性を定量化することが可能にする方法が提案されている。また，実データ実験において，提案法を用いてHIVウイルス薬に関連する遺伝子配列の突然変異の組み合わせを抽出し，得られた結果の統計的な信頼性を定量化することが可能となっている。

第二章では，機械学習の学習データに潜む外れ値（Outlier）が学習結果に与える影響度合いを自動的に制御し，学習結果を安定させるロバスト学習の新手法が提案されている．この研究のモチベーションは，近年，クラウドソーシングなどを利用して比較的容易にデータを取得することができる反面，それらのデータに潜むノイズにより学習結果が不安定になるという問題点を克服することである．提案法は，あるデータがどの程度異常であるかを自動的に判定しその影響度を調節することで学習結果を安定させることができる。影響度の調節はアニーリング（焼きまなし）法により実現されるが，提案法は無限に細かい粒度で（連続的に）温度パラメータを調節することが計算上可能であり，温度パラメータを離散的に調節する従来法と比較して高い汎化性能が得られることが数値実験で示されている．また，本研究では提案アルゴリズムによって得られる解の性質を理論的な側面から解析し，アルゴリズムの解釈性の向上が可能となることが示されている。

以上のように、本論文では機械学習による実データ分析において、信頼の高いロバストな方法を新たに提案したものであり、博士（工学）の学位を授与することが妥当である。